



OPEN ACCESS

EDITED BY

Constanza Cárdenas Carvajal,
Pontificia Universidad Católica de Valparaíso,
Chile

REVIEWED BY

Ganesh Panzade,
National Cancer Institute at Frederick (NIH),
United States
Akanksha Rajput,
University of California, San Diego,
United States

*CORRESPONDENCE

Clara Isabel Bermudez-Santana,
✉ cibermudezs@unal.edu.co

RECEIVED 13 September 2023

ACCEPTED 08 January 2024

PUBLISHED 31 January 2024

CITATION

Gutierrez-Diaz A, Hoffmann S,
Gallego-Gómez JC and Bermudez-Santana CI
(2024), Systematic computational hunting for
small RNAs derived from ncRNAs during
dengue virus infection in endothelial HMEC-
1 cells.
Front. Bioinform. 4:1293412.
doi: 10.3389/fbinf.2024.1293412

COPYRIGHT

© 2024 Gutierrez-Diaz, Hoffmann, Gallego-
Gómez and Bermudez-Santana. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](#). The use, distribution or reproduction in
other forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Systematic computational hunting for small RNAs derived from ncRNAs during dengue virus infection in endothelial HMEC-1 cells

Aimer Gutierrez-Diaz¹, Steve Hoffmann²,
Juan Carlos Gallego-Gómez³ and
Clara Isabel Bermudez-Santana^{1*}

¹Grupo Rnómica Teórica y Computacional, Departamento de Biología, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia, ²Faculty of Biosciences, Leibniz Institute on Aging—Fritz Lipmann Institute (FLI), Friedrich Schiller University Jena, Jena, Germany, ³Molecular and Translational Medicine Group, Medicine Faculty Universidad de Antioquia, Medellín, Colombia

In recent years, a population of small RNA fragments derived from non-coding RNAs (sfd-RNAs) has gained significant interest due to its functional and structural resemblance to miRNAs, adding another level of complexity to our comprehension of small-RNA-mediated gene regulation. Despite this, scientists need more tools to test the differential expression of sfd-RNAs since the current methods to detect miRNAs may not be directly applied to them. The primary reasons are the lack of accurate small RNA and ncRNA annotation, the multi-mapping read (MMR) placement, and the multicopy nature of ncRNAs in the human genome. To solve these issues, a methodology that allows the detection of differentially expressed sfd-RNAs, including canonical miRNAs, by using an integrated copy-number-corrected ncRNA annotation was implemented. This approach was coupled with sixteen different computational strategies composed of combinations of four aligners and four normalization methods to provide a rank-order of prediction for each differentially expressed sfd-RNA. By systematically addressing the three main problems, we could detect differentially expressed miRNAs and sfd-RNAs in dengue virus-infected human dermal microvascular endothelial cells. Although more biological evaluations are required, two molecular targets of the hsa-mir-103a and hsa-mir-494 (CDK5 and PI3/AKT) appear relevant for dengue virus (DENV) infections. Here, we performed a comprehensive annotation and differential expression analysis, which can be applied in other studies addressing the role of small fragment RNA populations derived from ncRNAs in virus infection.

KEYWORDS

differential expression, small RNASeq, ncRNAs, multimapping problem, dengue virus, tRNA-derived fragments

Introduction

Non-coding RNAs (ncRNAs) are functional RNA molecules involved in many cellular processes. During the last 20 years, ncRNAs have repeatedly been reported as sources of small RNAs. For example, alternative RNA processing steps for tRNAs give rise to smaller functional RNAs (Cole et al., 2009; Lee et al., 2009; Lyons et al., 2018).

Specifically, studies have found small RNA fragments between 20 and 30 nt in length mainly derived from tRNAs with a functional resemblance to the Argonaute-miRNA gene silencing pathway (Cole et al., 2009; Kumar et al., 2014; Kuscu et al., 2018). These processes can be observed across the tree of life. For instance, specific cleavage and processing of tRNAs have also been observed in the fungus *Aspergillus fumigatus* (Christoph et al., 2008) or human small RNA-seq libraries (Kawaji et al., 2008). In addition, recent surveys have also recognized that viruses employ a mechanism to suppress antiviral responses that involve the expression of RNA fragments derived from the host transfer RNA, which were in the past the so-called tRNA-derived fragments, or tRFs (Ivanov, 2015). For instance, respiratory syncytial virus (RSV)-induced tRFs derived from the 5-end of mature tRNAs decoding GlyCCC, LysCTT, and CysGCA (named tRF5-GlyCCC, tRF5-LysCTT, and tRF5-CysGCA, respectively) (Zhou et al., 2017) can target specific genes, e.g., SYNE-2, MIP-1 β , and IP-10 (Chen and Bao, 2020). tRFs have also been associated with antibody response to bovine leukemia virus (BLV) in Holstein cattle (Taxis et al., 2018). Finally, regulating La/SSB-dependent viral gene expression by pre-tRNA 3' trailer-derived small RNAs has been experimentally tested under various RNA virus infections (Cho et al., 2019). Apart from tRNAs, other ncRNAs linked to small RNA expression are precursor-microRNA (pre-miRNA) hairpins that give rise to additional micro RNA offsetRNAs (moRNAs) in *Ciona intestinalis* (Shi et al., 2009) and in the human brain (Langenberger et al., 2009). Small nucleolar RNAs (snoRNAs) have also been reported as a source for specific miRNA-like short RNAs (Ender et al., 2008; Saraiya and Wang, 2008; Taft et al., 2009) as well as for the reported data for vault RNAs (Persson et al., 2009; Stadler et al., 2009).

Overall, small RNA-mediated gene regulation may be even more complicated than previously believed (Christoph et al., 2008), e.g., affecting the human brain (Kawaji et al., 2008) or potentially influencing human diseases (Rashad et al., 2020). Therefore, the detection of new small RNA *loci* derived from annotated or un-annotated ncRNAs is a relevant problem to many fields in life sciences, in particular for inferring new ncRNA-disease connections, e.g., in cancer research (Zhu et al., 2020). Though next-generation sequencing (NGS) protocols have enabled us to discover expression profiles for many coding genes, those approaches may need to be more straightforwardly applied to ncRNAs. Therefore, to comprehensively study the expression profile of ncRNAs, several requirements must be addressed to minimize, for example, the impact of data analysis methods on the recovery of the accuracy of miRNA profiles (Tam et al., 2015). A more fine-grained transcriptome data analysis to identify alternative small RNA processing patterns could provide the opportunity to detect new biological pathways for small RNA processing and its associated functions. However, to accurately detect sfd-RNAs, it is necessary to improve the accuracy of detection of small RNA profiles.

This study addressed the problems relevant to small RNA detection derived from NGS data preprocessing and ncRNA annotations that come partly from the ncRNA copy-number

expansions in genome evolution. Accurate annotation of ncRNAs and the effect of the multi-mapping read (MMR) placement problem on differential expression analysis. We tackle the problem of assigning the correct expression levels of ncRNAs because their expansion during genome evolution lets them be presented as multicopy genes, as seen in the case of tRNAs in many eukaryotes (Bermudez-Santana et al., 2010; Velandia-Huerto et al., 2016). In addition, other kinds of repeated and non-coding regions exist in eukaryotic chromosomes—centromeres, pericentromeric regions, and telomeric regions— and are interspersed between them. However, establishing their function is currently being researched since authors recognize the need to improve methods to study the biology and methodology of small RNA in wet-lab experiments and silico approaches. However, new advances are found in the study by Louzada et al. (2020) and Lopes et al. (2023).

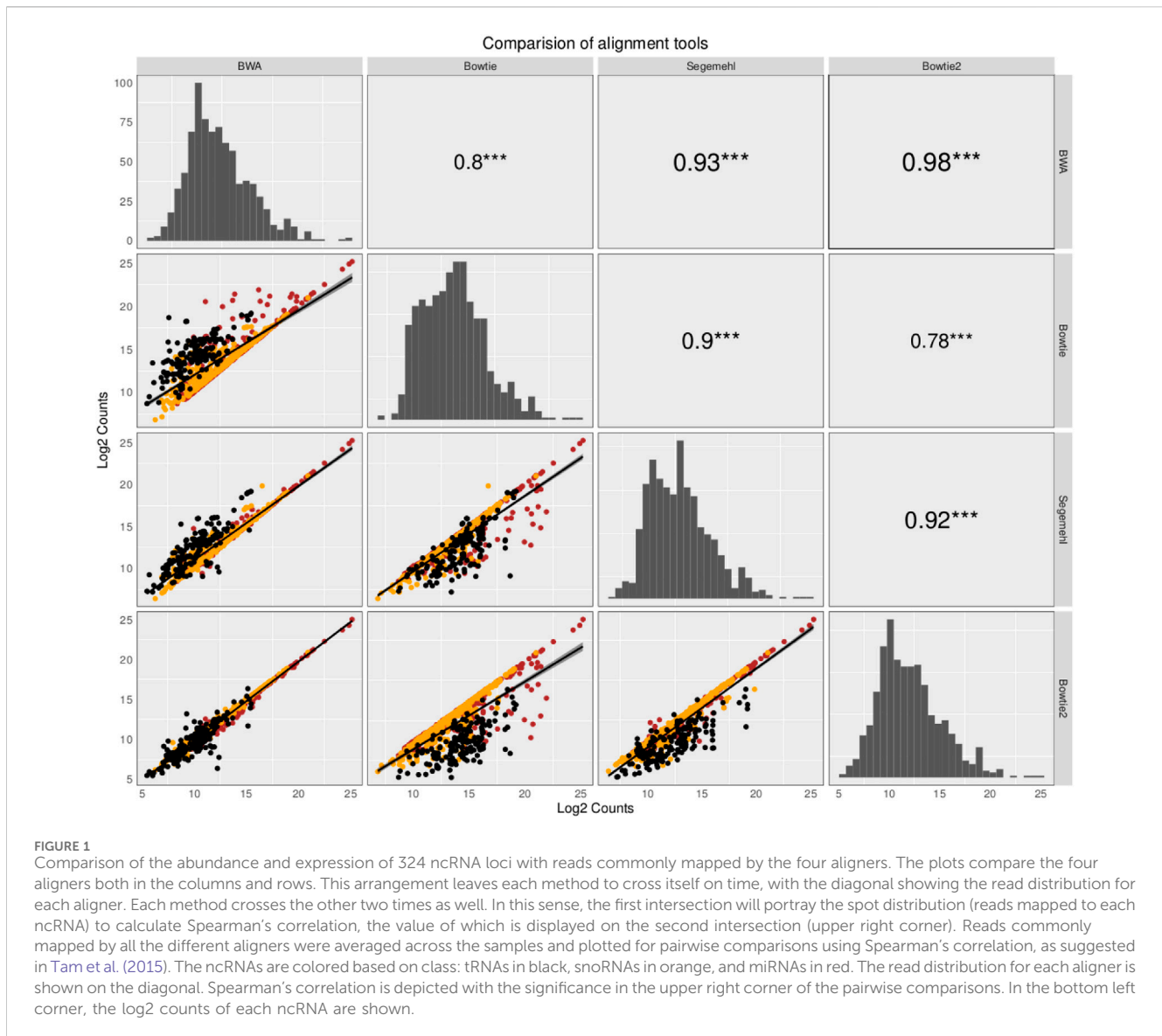
We developed a solution to these problems by detecting several differentially expressed sfd-RNAs that might activate in the host by DENV infection. It remains to be evaluated if they have a unique role in non-canonical RNAi pathways. Nonetheless, during the last decade, the importance of ncRNAs in the regulation of viral replication, viral persistence, host immune evasion, and cellular transformation has increasingly gained attention, e.g., in the case of viral non-coding RNAs (Tycowski et al., 2015; Ivanov, 2015; Nunes et al., 2020). Recent results from the study by our group suggest a potential biological role for some sfd-RNA candidates. For some miRNAs on the starting point, consolidation, and finishing of the viral cycle of DENV in cell cultures (Roa-Linares, 2019; Cuartas-Lopez et al., 2018), assessing the biological function for the majority of the sfd-RNAs detected here remains a future challenge.

Results

ncRNA unified annotation, without copy-number limitations

The annotation of families of miRNAs, tRNAs, and snoRNAs in public databases remains problematic (Kozomara and Griffiths-Jones, 2013; Telonis et al., 2015; Jorjani et al., 2016) and creates ambiguities in the assignment of sfd-RNAs to their sources (Schopman et al., 2010). In addition, the multicopy nature of ncRNAs aggravates the problem of assigning sources for sfd-RNAs and establishing their new biological function. However, few *loci* showed strong evidence for a natural role (Hoeppner et al., 2018).

Despite experimental constraints to detecting and testing the expression for ncRNAs in multiple copies, we faced the problem of providing a measure of specificity for an ncRNA to be able to code for specific small RNAs. For instance, categories of homologous ncRNA sequences were defined to solve the conflict of unsolved annotation for ncRNAs. This approach also helped clean identifiers approved by the HGNC. We set some categories to help simultaneously classify and integrate ncRNA annotation and avoid species-specific annotation changes due to the curation; for instance, we used ncRNA annotation of conserved sequences in primates. Additionally, we included all ncRNA categories to classify



and integrate the ncRNAs even in cases in which an ncRNA fit multiple class annotations as was seen for the *loci* coding the hsa-mir-768 and the snoRNA HBII-239 (Hüttenhofer et al., 2001). Furthermore, a method to reduce the overrepresented identical paralogs of ncRNAs was included (see methods: *Consensus of multicopies of ncRNA loci*). Our strategy to solve the conflicting annotations and the integration of the ncRNA gene features allowed us to find several sfd-RNAs annotated under other ncRNA categories, such as piRNAs and snoRNAs. Notoriously, those piRNAs resemble the coordinates of recent significant discoveries such as the piR-hsa-23289, which have the same genome coordinates defined for the tRF experimental validated tRF-5'-GluCTC (Wang et al., 2013).

Effect of multi-mapping reads on ncRNAs

The small RNA sequencing (RNA-seq) products analyzed here are derived from two fractions of RNA of sizes ranging from

15 to 30 nt (small-sized fraction) and from 30 to 200 nt (medium-sized fraction). Fractions were size-purified from total RNA isolated from DENV2-infected endothelial (HMEC-1) cells and non-infected HMEC-1 cells (mock cells-controls) at different time points. Reads were aligned to 4,320 ncRNAs by all four aligners. Then, two comparisons were made to identify the bias generated by each aligner method. Those results have been portrayed in Figures 1, 2. The plots compare the four aligners, both in columns and rows. This arrangement leaves each method to cross itself on time, with the diagonal showing the read distribution for each aligner. Each method crosses the other two times as well. In this sense, the first intersection will portray the spot distribution (reads mapped to each ncRNA) to calculate Spearman's correlation, the value of which is displayed on the second intersection (upper right corner).

In the first comparison, reads commonly mapped by all the different aligners were averaged across the samples and plotted for pairwise comparisons using Spearman's correlation as suggested in Tam et al. (2015). In Figure 1, which contains information from that

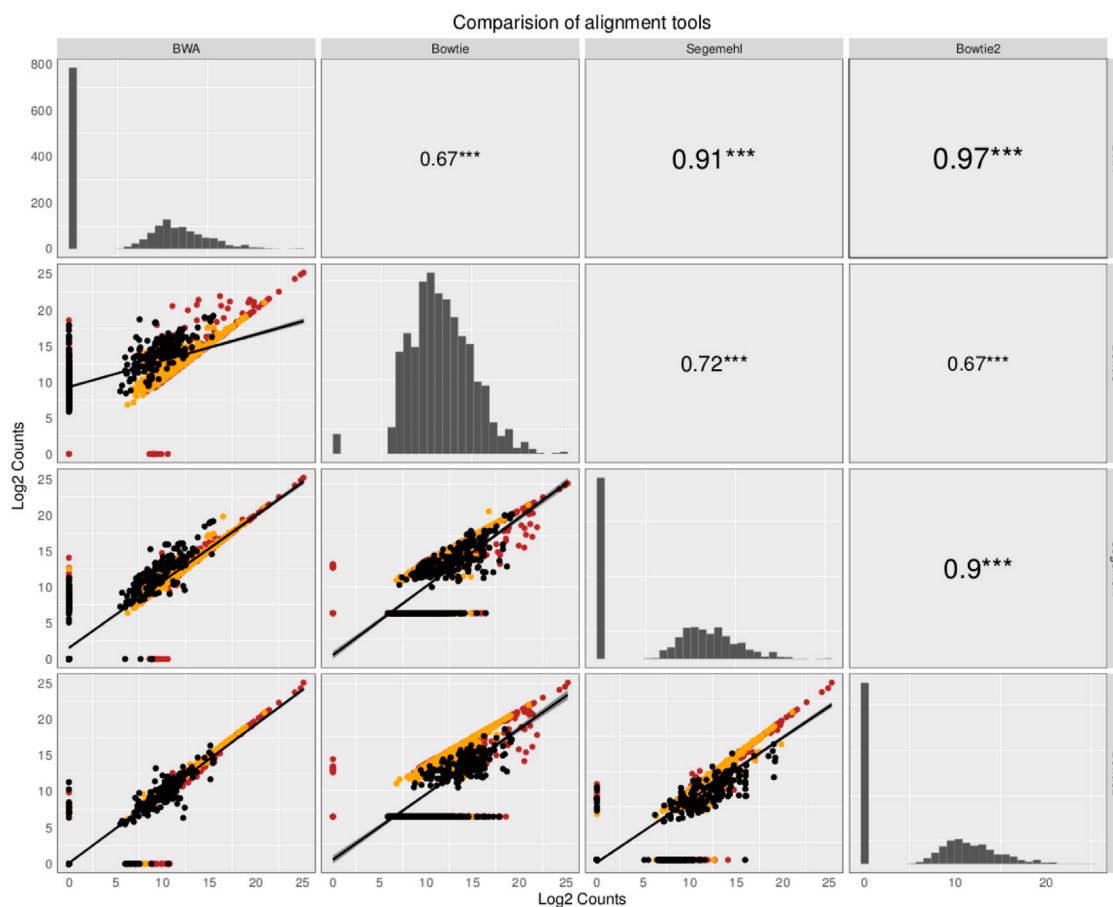


FIGURE 2

Comparison of the abundance and expression of the total number of ncRNA with read alignments. The plots compare the four aligners both in the columns and rows. This arrangement leaves each method to cross itself on time, with the diagonal showing the read distribution for each aligner. Each method crosses the other two times as well. In this sense, the first intersection will portray the spot distribution (reads mapped to each ncRNA) to calculate Spearman's correlation, the value of which is displayed on the second intersection (upper right corner). The ncRNAs are colored based on class: tRNAs in black, snoRNAs in orange, and miRNAs in red. Here, all the mapped reads are included, regardless of whether all the aligners commonly map the reads. Although some aligners have a high correlation value, due to the similarities in the alignment algorithm, each mapper produces a new particular output, giving evidence that the aligner methods influence the expected results in research of small ncRNAs.

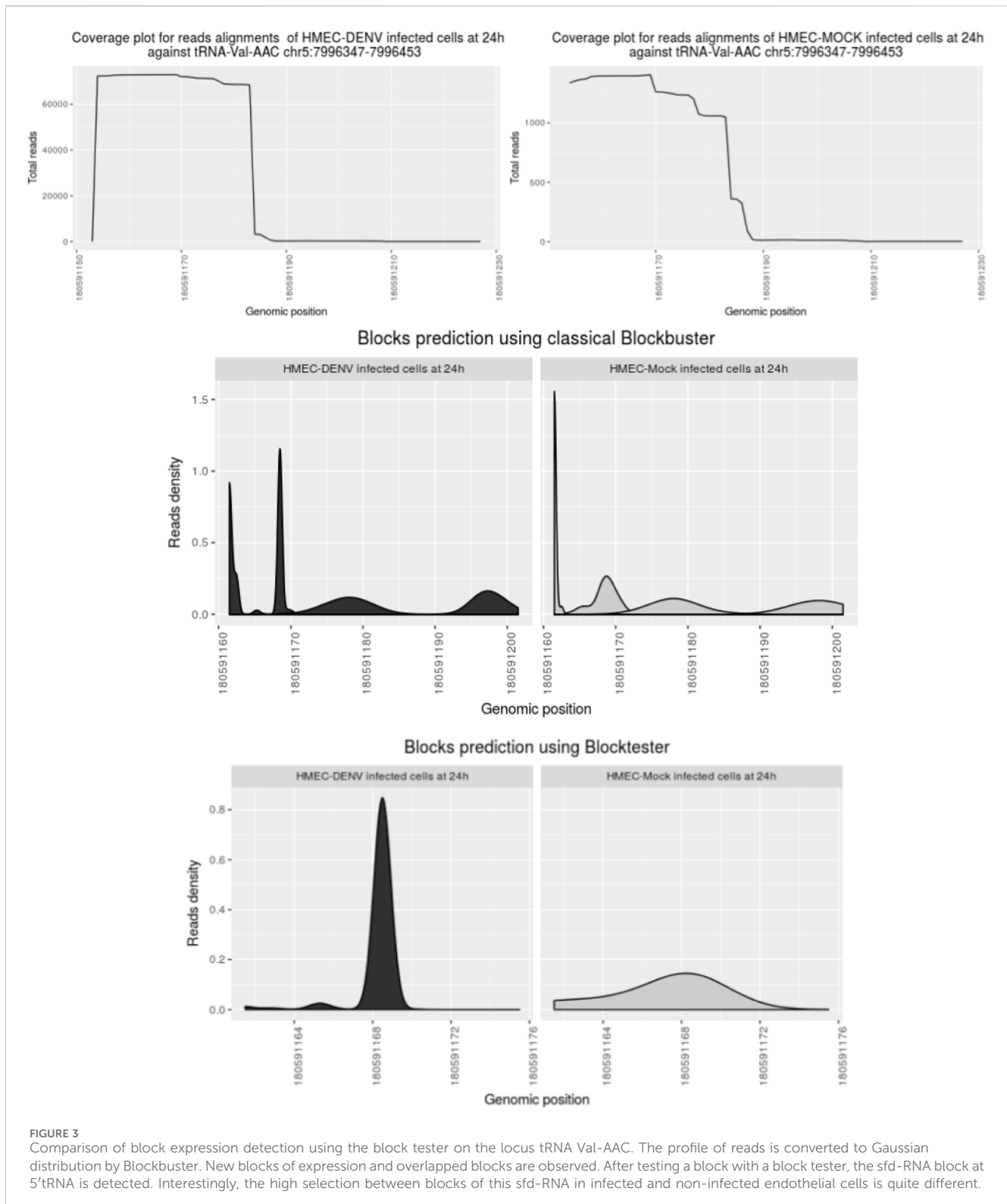
comparison, the value of Spearman's correlation coefficient is over 0.9 for all the pairwise comparisons, except for the comparison with Bowtie, but the correlation between Bowtie and segemehl is 0.9, which can handle MMR reads and recover the highest number of reads aligned to ncRNAs. However, the restriction to working only with reads mapped by all the aligners reduces the specific advantages that each mapping algorithm has. While BWA aligns MMRs with a random behavior to one genomic location, Bowtie, Bowtie2, and segemehl aligners generally avoid a pseudorandom distribution of reads among repetitions.

In contrast, in the second comparison, when all the ncRNAs (including all the mapped reads, regardless of whether all the aligners commonly map the reads) are included to compare the aligners, Spearman's correlation behaves similarly between all the aligner values, but it changes dramatically in correlations mainly in paired comparisons between other aligners and Bowtie (Spearman's correlation coefficient varies from 0.67 to 0.72); see the behavior in Figure 2. Interestingly, several ncRNAs explicitly detected by this aligner were also detected as differentially expressed (see the

following section). Intriguingly, noticing how the read distribution changes for each aligner on the diagonal arrangement is essential. Furthermore, we noted that ncRNAs with more copies, such as tRNAs, have an aligner-dependent behavior due to *sfd*-RNAs processed from these transcripts, mainly detected as differentially expressed by aligners with tolerance to MMRs.

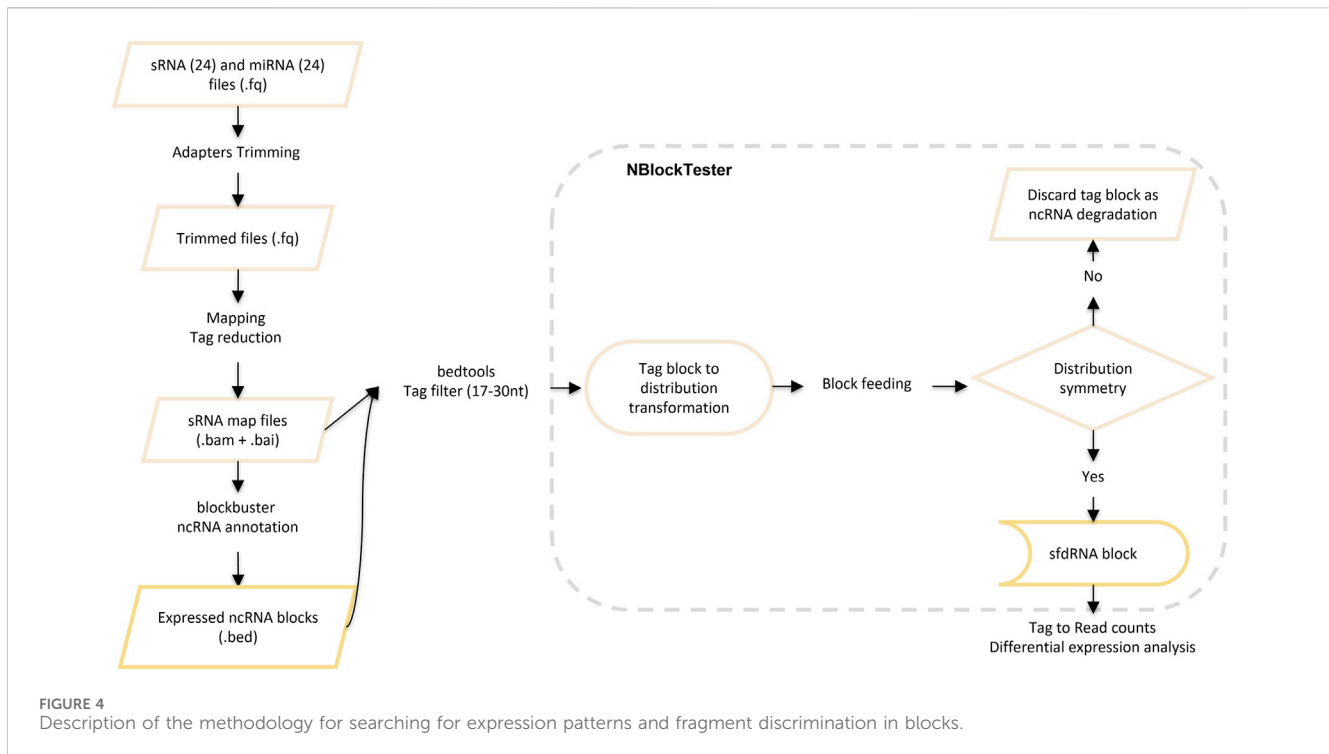
Differential expression detection of *sfd*-RNAs and miRNAs

After mapping, we detected ncRNAs with complex profiles of expression. Those patterns exhibit nested blocks built of a pile of reads that arise from two different but adjacent positions in the same locus. Figure 3 illustrates a typical example showing such patterns. First, Blockbuster (Langenberger et al., 2009) was used to detect the global expression and patterns for each ncRNA locus once the ncRNA *loci* or classes were defined. Then, a further step was



carried out to tune the detection of expression profiles of *sfd*-RNAs embedded in a complex distribution of reads. In this new step, a test over each block of the entire mapped region is done independently of the accumulation of dispersed reads (see methods: *Searching for expression patterns and fragment discrimination in blocks*. See

Figure 4 for the summary of the explanation of the procedure). This new step uncovered several small RNAs on *loci*, which had not previously been associated with production of small RNAs. Then, small- and medium-sized fractions were contrasted to discriminate between degradation fragments or non-fragmented



ncRNAs and induced sfd-RNAs by dengue virus infection. After comparing and measuring the basal expression of ncRNAs in uninfected (mock cells—controls) and infected cells in the medium-sized fraction, the small-sized fraction was used to detect the differential expression by contrasting block expressions against the controls simultaneously in both fractions. Putative sfd-RNAs in blocks were deemed differentially expressed if the expression level changes were statistically significant under dengue virus infection at any time post-infection, but none in any control.

The expression patterns for the selected differentially expressed sfd-RNAs are represented in a heatmap. The global profile for the candidates with the best statistical values and the best rank order of prediction for each differentially expressed sfd-RNA, shown in Figure 5. Apart from the miRNAs hsa-mir-103a and hsa-mir-494, tRNA-derived sfd-RNAs, e.g., tRF-3'ProTGG, tRF-3'ProAGG, tRF-3' GluCTC, and tRF-5'ValACC, were also ranked with a consistent score of reproducibility of 1 for all the sixteen strategies, i.e., identified as differentially expressed by the four aligners and normalization methods. The summary of statistics of these results is shown in Table 1.

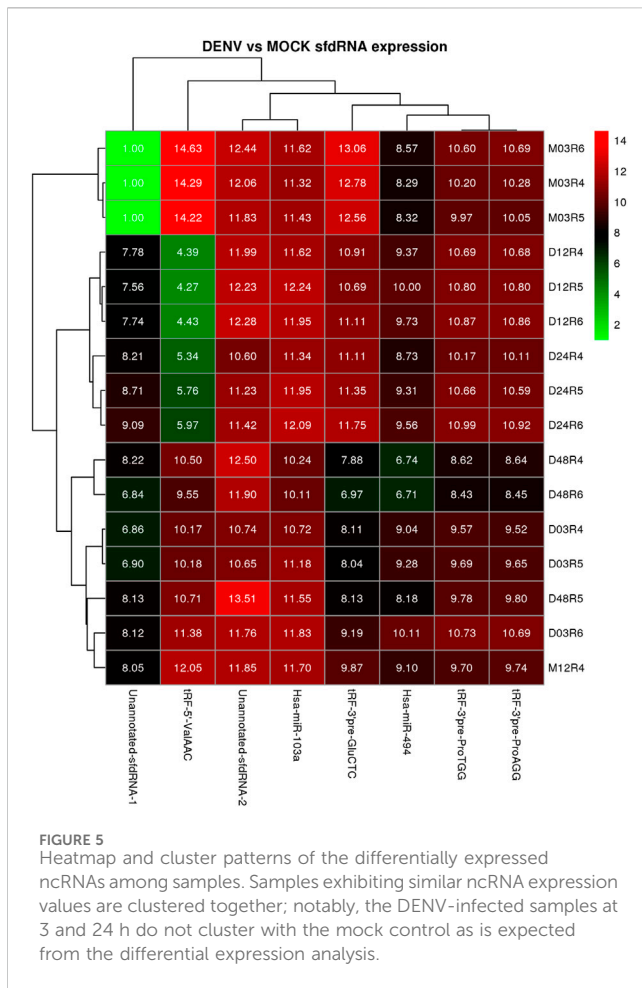
Discussion

Interestingly, after dealing with the multi-mapped read problem and checking the mapping and normalization methods, we found intricate patterns of piles of mapped reads. As the first impression, those signals presumably indicate a pattern of reads that were generally considered random degradation products previously. While analyzing those intricate patterns, the automated detection

of sfd-RNAs became another arduous task. Although tools like FlaiMapper (Hoogstrate et al., 2014) attempt to address the challenge of handling nested blocks, the problem is only sometimes solved. In short, because a minimum number of identical read starts and ends is required to annotate a block, piles of reads indicative of sfd-RNAs often need to be found. Our approach dealt with this problem, helped us solve complex read assignments, and facilitated differential expression analysis.

The proposed improvements help determine sets of up- and downregulated sfd-RNAs in particular experimental conditions. The expression analysis of hsa-mir-103a and hsa-mir-494 48 h after dengue virus infection is an essential example of this development. Nevertheless, tRFs were not always predicted simultaneously using the four aligners. Still, the advantage of using more than one strategy was highly desirable to detect them since a rank-order of prediction of 1 was observed for tRFs tRF-3'ProTGG, tRF-3'ProAGG, tRF-3' GluCTC, and tRF-5'ValACC. Remarkably, tRF-5'ValACC got a fold change of 10.5 from 12 to 48 h post-infection. Process validation should be defined to document the evidence of this change.

The expression profile of the top eight differentially expressed sfd-RNAs in endothelial cells infected by DENV2 gave the best predictions for two miRNAs (hsa-mir-103a and hsa-mir-494) as differentially expressed between cell cultures. These results are compatible with those of our previous published work. Moncini et al. (2011) found out that the miR-103 is implied in regulating CDK5R1, a transcript related to cellular migration. In addition, Roa-Linares (2019) have reported that the loss of function of CDK5 causes cytoskeleton reorganization and a decrease in dengue virus infection. Additionally, miR-494 promotes PI3K/AKT pathway hyperactivation in human hepatocellular



remodeling of microfilaments of actin after dengue virus infection on hepatoma cell cultures (Cuartas-Lopez et al., 2018). In addition, the role of small RNAs in dengue virus 2-infected Aedes mosquito cells has also revealed viral piRNAs and novel host miRNA expression that changed substantially upon DENV2 infection (Miesen et al., 2016). In addition, sfd-RNAs derived from tRNAs, tRF-3'ProTGG, tRF-3'ProAGG, tRF-3'GluCTC, and tRF-5'ValAAC, were also ranked with a consistent score of reproducibility for all the sixteen strategies. These results are in agreement with the fact that for some tRFs, there is a substantial increase in their abundance in chronic viral infections of the liver caused by hepatitis B virus (HBV) or hepatitis C virus (HCV) (Selitsky et al., 2015). Similar tRFs expressed under other experimental conditions are deposited at the tRFdb (Kumar et al., 2015; tRFdb, 2020).

Our exhaustive analysis of the expression of small RNAs under DENV infection led us to detect candidates of sfd-RNAs related to other sfd-RNAs that were validated to be products of ncRNAs (Fort et al., 2022) or specifically of tRNAs (Cole et al., 2009) in viral infection conditions (Zhou et al., 2017; Taxis et al., 2018; Cho et al., 2019; Chen and Bao, 2020) or under other biological conditions that affect their cleavage and processing (Christoph et al., 2008; Kawaji et al., 2008; Hoogstrate et al., 2014). Furthermore, our work proposes new candidates whose differential expression must be experimentally tested, particularly those suited to investigate whether our candidates indeed are regulated under dengue virus infection. This study has potential limitations. Our model is based on the annotation derived from hg19. They are, therefore, subject to biases and confounding factors that may have influenced our model estimates. However, the coordinates of differentially expressed sfd-RNA candidates were all corrected and assigned to the coordinates of the human GRCh38.p14 version with the accepted HGCN names of potential sources of sfd-RNAs (that information is available in Supplementary Tables S1–S4 including the coordinates of the sfd-RNAs, the nearest annotated source, and the distance to it).

carcinoma progression (Lin et al., 2018). Recently, we found evidence from several experiments that PI3/AKT pathway activation controls the small RhoGTPases and ulterior

TABLE 1 Summary statistics of representative sfd-RNAs and miRNAs differentially expressed in DENV2-infected HMEC-1 cells. FC: fold change; CPM: count per million; LR: likelihood ratio statistics; FDR false discovery rate; Aligner and NM columns display the computational method where the sfdRNA has the highest LR value where Aligner represents the aligner tool and NM to the normalization method. RA is the fraction of reproducibility of the aligner concerning the other aligners, and RN is the fraction of reproducibility of all normalization methods. The abbreviations CPM, UQS, TMM, and RLE represent count per million, upper quartile scaling, trimmed Mean of M, and the relative log expression, respectively. The complete set of identified sfd-RNAs as DE is available in Supplementary Tables S1–S4.

sfd-RNA source	GRCh38.p14 position	hpi	logFC	logCPM	LR	p-value	FDR	Aligner	RA	NM	RN
POLA1	X:24970573–24970603	3	10.5	6.4	23.62	1.0E-6	8.0E-4	Bowtie2	0.5	UPQ	1
Not assigned	chr17:53088644–53088661	24	-3.16	10.51	12.58	4.0E-4	3.0E-2	Bowtie2	0.5	CPM	1
MIR103A2	chr20:3917541–3917563	48	4.33	10.3	14.16	2.0E-4	2.0E-2	segemehl	0.25	UPQ	1
MIR494	chr14:101023375–101023396	48	6.21	7.92	15.16	1.0E-4	1.0E-2	segemehl	0.25	RLE	1
tRF-3' tRNA-Pro-TGG-3-1	chr5:181188895–181188933	24	1.77	8.27	15.21	1.0E-4	1.0E-2	Bowtie2	0.25	RLE	1
tRF-3'tRNA-Pro-AGG-2-5	chr14:20609377–20609415	24	1.79	8.22	15.6	8.0E-5	1.0E-2	Bowtie2	0.25	RLE	1
tRF-3'tRNA-Glu-CTC-1-2 (pre-GluCTC)	chr1:161447258–161447307	48vs24	9.96	11.48	25.86	4.00E-07	0.0002	BWA	0.25	TMM	1
tRF-5' TRV-AAC1-2 (tRNA-ValAAC)	chr5:181164154–181164183	48vs12	10.06	12.34	21.94	3.0E-6	3.0E-3	BWA	0.25	RLE	1

Conclusion

Our approach allowed us to quantify simultaneously the differential expression of miRNAs and sfd-RNAs by an exhaustive analysis of small RNA-seq data derived from DENV-2-infected endothelial cells. Spearman's correlation coefficient value measured among the four types of aligners may suggest that the alignment method has a considerably more significant impact on the recovery of ncRNA abundance profiles than the library normalization method. The new ranking score introduced here helps prioritize subsequent experimental validation of candidate sfd-RNAs since it is based on the results of sixteen combinations of aligners and normalization methods. Our computational procedure allowed us to detect un-annotated *loci* acting as sources of small RNAs. We found significant ranking scores for the human miRNAs hsa-mir-103a and hsa-mir-494 during DENV-2 infection of endothelial HMEC-1 cell cultures. Our findings indicate that choosing an aligner would significantly impact the detection of the candidate sources of sfd-RNAs. Furthermore, re-inspecting expression profiles resulted in an accurate count of reads essential for differential expression analysis. Additionally, although the role of these sfd-RNAs during DENV infection remains to be validated, previously published findings provided strong evidence for potential cellular, molecular targets (CDK5 and PI3/AKT) of the miRNAs hsa-mir-103a and hsa-mir-494. The fact that tRFs are changed in various viral infections argues for including them in future studies.

Methods

Total RNA extraction and small RNA library preparation and sequencing

Cell lines and synchronized DENV2 infections are detailed in Alvarez-Diaz et al. (2019). DENV2-infected and uninfected HMEC-1 cells were considered for RNA extraction at four time points (3, 12, 24, and 48 h, each with three replicates). RNA of sizes ranging from 15 to 30 nt (small-sized fraction) and from 30 to 200 nt (medium-sized fraction) was extracted. A total of 48 libraries (24 for each size) were prepared and sequenced on the Illumina NextSeq 500 system by Exiqon facilities under the same protocol, as described in Alvarez-Diaz et al. (2019). The raw data of libraries are deposited at the NCBI, BioProject ID: PRJNA1018251.

Small RNA-seq analysis

Consensus of multicopies of ncRNA loci

Sequences of miRNAs, tRNAs, and snoRNAs were obtained from UCSC (hg19), Rfam, snoRNABase download page, and tRNAscan-SE predictions. Subsequently, all sequences were mapped with BLASTN to the human genome (hg19) to search for multiple hits in the genome. Then, to define a unique set of homologous *loci*, the consensus of sequences was calculated using BLASTclust (Dondoshansky and Wolf, 2002) and bedtools (Quinlan and Hall, 2010). UCSC tracks included the miRBase version 21

(Griffiths-Jones et al., 2006) (a total of 1,871 pre-miRNAs), UCSC (a total of 939 pre-miRNAs) (Karolchik et al., 2004), and Rfam version 11.0 (a total of 1,229 pre-miRNAs) (Griffiths-Jones et al., 2005). Finally, we used 2,075 coordinates for 1,985 different miRNAs for further analysis. Following the same procedure, 937 snoRNA genomic *loci* were finally solved, starting from 858 genomic coordinates from Rfam and 486 from UCSC. Finally, previously detected human nuclear chromosomes of multi-sequences highly similar to human mitochondrial tRNAs (tRNA-lookalikes) (Telonis et al., 2015) increased the number of tRNA genomic positions to 1,129. Thus, we used the same procedure applied to miRNAs and snoRNAs to get a final set of 940 unique sequences for tRNAs used in further steps.

Raw data pre-processing and quality control

We evaluate the quality of the NGS data using FastQC (0.11.2). Low-quality reads with lengths less than 16 nt were dismissed. The over-representative 3' adapters reported were also trimmed to avoid the high false-positive rates and detect typical clippers when applied to small RNA libraries.

Mapping of reads

Four aligners, namely, BWA (Li and Durbin, 2009), Bowtie (Langmead et al., 2009), Bowtie2 (Langmead and Salzberg, 2012), and segemehl (Hoffmann et al., 2009), were used to map reads to the human genome reference hg19. Tags were used to represent equal reads to reduce the mapping computational time.

Parameters used in this analysis are described in Tam et al. (2015). SAM outputs were converted to BAM format using SAMtools (Li et al., 2009). BAM files were used as the inputs to search for block expressions. Reads were considered a favorable map if they matched with the consensus ncRNAs on the same strand and inside the locus boundaries.

Searching for expression patterns and fragment discrimination in blocks

In the first step, blocks of mapped reads were calculated by Blockbuster (Langenberger et al., 2009). In the second step, the BAM format of mapped reads in consensus ncRNAs was filtered by size to filter the expected size of sfd-RNAs candidates (17–30 nt long). Then, a routine to tune reads assigned to block candidates to sfdRNAs was implemented (<https://github.com/AimerGDiaz/NBlockTester>). NBlockTester starts calculating the midpoint of each read u by $\mu_u = (b_u + a_u)/2$, where a_u and b_u correspond to the start and end positions, respectively, and variance σ_u^2 is defined by $\sigma_u = (b_u - a_u)/2$. To avoid misclassified reads of sfd-RNAs in a block, we introduced a symmetry criterion to tune the classification of reads in blocks. Here, we started calculating the minimum midpoint from a set of midpoints, i.e., a set of reads; the initial symmetry range sr is defined by $sr = \mu_{u_0} + \sigma_{u_0} * 1/2$. Then, μ is calculated for the next read and compared with sr . If $\mu < sr$, this read will be a candidate for this block. In summary, for all the reads clustered in a block in a candidate locus, the minimum midpoint value corresponds to the lowest mid coordinate of all the reads mapped to it. Then, for the growing block, the mean and the variance will be replaced by the arithmetic average $\hat{x} = \frac{1}{n} \sum_{i=1}^n x_i$ and the standard deviation $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{x})^2}$, where x_i represent the starting and end coordinates a_{u_i} , b_{u_i} , $a_{u_{i+1}}$,

$b_{ui+1}, \dots, a_{u_n}, b_{u_n}$ i.e., calculated on each step with the accumulated reads of the growing block. The full explanation of the procedure is summarized in Figure 4.

Counting reads in blocks

Once blocks of read candidates to represent the expression of sfd-RNAs were set, we traced the reads to identify those mapped to multiple *loci*. Mapped reads from each ncRNA locus were retrieved with SAMtools using the coordinates of the ncRNA locus (independently of the annotation source). By the procedure explained in the *Consensus of multicopy of the ncRNA loci procedure*, those mapped reads were merged on consensus ncRNA categories. Then, to solve the MMR problem, a filter was applied to the predicted read blocks for each Bowtie and segemehl aligner, reducing identical multi-mapped reads to a unique representative read. However, in the case of the aligners Bowtie2 and BWA, which map pseudo-randomly on multiple homologous targets, the merged outcome will be used in the subsequent analysis.

Discriminating between degradation fragments or non-fragmented ncRNAs and induced sfd-RNAs by dengue virus infection

Expression levels from the small- and medium-sized fractions were compared between infected and uninfected cell cultures at each time point described above. Then, to discard functional subproducts (putative sfd-RNAs) induced by dengue virus infection of degradation fragments or non-fragmented ncRNAs, we first compared and contrasted the basal expression of ncRNAs in uninfected (mock cells) with the expression of ncRNAs from infected cells using the medium-sized fraction. Then, in the second step, the small-sized fraction was used to detect differential expression by simultaneously contrasting the block expression against the controls in both fractions. So a block (putative sfd-RNAs) was declared differentially expressed if a change in read counts or expression levels was statistically found any time post-infection, but none in any control.

Differential expression analysis

Differential expression analyses were performed in the R statistical environment (v3.2.0) using edgeR (v3.2.0) software (Bioconductor v2.12 repository). DENV2-infected and uninfected HMEC-1 cells at 3, 12, 24, and 48 h (with three replicates) were considered factors in constructing the design matrix used as input to estimate dispersions according to the Cox-Reid profile-adjusted likelihood (CR) method. We used four normalization methods, i.e., counts per million (CPM), upper quartile scaling (UQS), trimmed mean of M (TMM), and relative log expression (RLE), for the differential expression analyses following the indications of Tam et al. (2015). The read count in the design matrix was fitted to a generalized linear model (GLM). Fold changes greater than 1.0 or lower than -1.0 and FDR less than 0.05 were used as thresholds to

select the top set of differentially expressed sfd-RNA candidates. Then, our approach, coupled with four aligners and four normalization methods, lets us define sixteen different computational strategies. Then, a rank order of prediction for each sfd-RNA is measured as 1 for the sixteen computational strategies that predicted it as differentially expressed. Coordinates of hg19 and accepted HGCN names of potential sources of sfd-RNAs, which were statistically significant, were updated to the Genome Reference Consortium Human GRCh38.p14 (GCA_000001405.29) and UCSC Genome Browser assembly ID: hg38 using bigBedToBed and LiftOver available at UCSC (consulted on 29 November 2023).

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: (Bioproject Accession Number: PRJNA1018251).

Ethics statement

Ethical approval was not required for the studies on humans in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used. Ethical approval was not required for the studies on animals in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used.

Author contributions

AG-D: conceptualization, data curation, formal analysis, methodology, software, visualization, writing—original draft, and writing—review and editing. SH: conceptualization, investigation, methodology, software, supervision, and writing—review and editing. JG-G: conceptualization, funding acquisition, investigation, project administration, resources, supervision, and writing—review and editing. CB-S: conceptualization, funding acquisition, investigation, methodology, project administration, supervision, writing—original draft, and writing—review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The wet-lab experiments were funded by the Ministry of Science, Technology, and Innovation—Minciencias grant 111584466951, and CODI-UdeA 2020-34137. Data analysis was carried out using the equipment donated by the German Academic Exchange Service-DAAD and the Faculty of Science at the Universidad Nacional de Colombia. The first author was supported by the

scholarship graduate program at the Universidad Nacional de Colombia, Sede Bogotá Resolución 498-2015, Acta 13 23rd October 2015.

Acknowledgments

We appreciate the members of the Molecular and Translational Medicine Group and Grupo Rnomica Teórica y Computacional for their constructive discussions.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbinf.2024.1293412/full#supplementary-material>

SUPPLEMENTARY TABLE S1

Identified DE of sfd-RNA of (small-sized fraction) GRCh38.p14 coordinates and HGNC names.

SUPPLEMENTARY TABLE S2

Identified DE of sfd-RNA of (small-sized fraction) GRCh38.p14 coordinates and non-assigned HGNC names.

SUPPLEMENTARY TABLE S3

Identified DE of sfd-RNA of (medium-sized fraction) GRCh38.p14 coordinates and HGNC names.

SUPPLEMENTARY TABLE S4

Identified DE of sfd-RNA of (medium-sized fraction) GRCh38.p14 coordinates and non-assigned HGNC names.

References

- Alvarez-Diaz, D. A., Gutierrez-Diaz, A. A., Orozco-Garcia, E., Puerta-Gonzalez, A., Bermudez-Santana, C. I., and Gallego-Gomez, J. C. (2019). Dengue virus potentially promotes migratory responses on endothelial cells by enhancing pro-migratory soluble factors and miRNAs. *Virus Res.* 259, 68–76. doi:10.1016/j.virusres.2018.10.018
- Bermudez-Santana, C., Attolini, C. S., Kirsten, T., Engelhardt, J., Prohaska, S. J., Steigle, S., et al. (2010). Genomic organization of eukaryotic tRNAs. *BMC Genomics* 11, 270. doi:10.1186/1471-2164-11-270
- Chen, Y., and Bao, X. (2020). Respiratory syncytial virus induces a functional tRNA-derived fragment to promote infection by targeting SYNE2. *J. Immunol.* 204 (1), 93.2. doi:10.4049/jimmunol.204.supp.93.2
- Cho, H., Lee, W., Kim, G. W., Lee, S. H., Moon, J. S., Kim, M., et al. (2019). Regulation of La/SSB-dependent viral gene expression by pre-tRNA 3' trailer-derived tRNA fragments. *Nucleic Acids Res.* 47 (18), 9888–9901. doi:10.1093/nar/gkz732
- Christoph, J., Rederstorff, M., Hertel, J., Stadler, P. F., Hofacker, I. L., Schrettl, M., et al. (2008). Small ncRNA transcriptome analysis from *Aspergillus fumigatus* suggests a novel mechanism for regulation of protein-synthesis. *Nucleic Acids Res.* 36, 2677–2689. doi:10.1093/nar/gkn123
- Cole, C., Sobala, A., Lu, C., Thatcher, S. R., Bowman, A., Brown, J. W., et al. (2009). Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *RNA* 15 (12), 2147–2160. doi:10.1261/rna.1738409
- Cuartas-Lopez, A. M., Hernandez-Cuellar, C. E., and Gallego-Gomez, J. C. (2018). Disentangling the role of PI3K/Akt, Rho GTPase and the actin cytoskeleton on dengue virus infection. *Virus Res.* 256, 153–165. doi:10.1016/j.virusres.2018.08.013
- Dondoshansky, I., and Wolf, Y. (2002). *Blastclust (ncbi software development toolkit)*. Bethesda, Md: NCBI.
- Ender, C., Krek, A., Friedlander, M. R., Beitzinger, M., Weinmann, L., Chen, W., et al. (2008). A human snoRNA with microRNA-like functions. *Mol. Cell* 32, 519–528. doi:10.1016/j.molcel.2008.10.017
- Fort, R. S., Chavez, S., Trinidad Barnech, J. M., Oliveira-Rizzo, C., Smircich, P., Sotelo-Silveira, J. R., et al. (2022). Current status of regulatory non-coding RNAs research in the tritryp. *Noncoding RNA* 8 (4), 54. doi:10.3390/ncrna8040054
- Griffiths-Jones, S., Grocock, R. J., Van Dongen, S., Bateman, A., and Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic acids Res.* 34 (Suppl. 1), D140–D144. doi:10.1093/nar/gkj112
- Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., and Bateman, A. (2005). Rfam: annotating non-coding RNAs in complete genomes. *Nucleic acids Res.* 33 (Suppl. 1), D121–D124. doi:10.1093/nar/gki081
- Hoepfner, M. P., Denisenko, E., Gardner, P. P., Schmeier, S., and Poole, A. M. (2018). An evaluation of function of multicopy noncoding RNAs in mammals using ENCODE/
- FANTOM data and comparative genomics. *Mol. Biol. Evol.* 35 (6), 1451–1462. doi:10.1093/molbev/msy046
- Hoffmann, S., Otto, C., Kurtz, S., Sharma, C. M., Khaitovich, P., Vogel, J., et al. (2009). Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput. Biol.* 5 (9), e1000502. doi:10.1371/journal.pcbi.1000502
- Hoogstrate, Y., Jenster, G., and Martens-Uzunova, E. S. (2014). FlaiMapper: computational annotation of small ncRNA-derived fragments using RNA-seq high-throughput data. *Bioinformatics* 31 (5), 665–673. doi:10.1093/bioinformatics/btu696
- Hüttenhofer, A., Kiefmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, J. P., et al. (2001). RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.* 20 (11), 2943–2953. doi:10.1093/emboj/20.11.2943
- Ivanov, P. (2015). Emerging roles of tRNA-derived fragments in viral infections: the case of respiratory syncytial virus. *Mol. Ther.* 23 (10), 1557–1558. doi:10.1038/mt.2015.161
- Jorjani, H., Kehr, S., Jedlinski, D. J., Gumienny, R., Hertel, J., Stadler, P. F., et al. (2016). An updated human snoRNAome. *Nucleic acids Res.* 44 (11), 5068–5082. doi:10.1093/nar/gkw386
- Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., et al. (2004). The UCSC Table Browser data retrieval tool. *Nucleic acids Res.* 32 (Suppl. 1), D493–D496. doi:10.1093/nar/gkh103
- Kawaji, H., Nakamura, M., Takahashi, Y., Sandelin, A., Katayama, S., Fukuda, S., et al. (2008). Hidden layers of human small RNAs. *BMC Genomics* 9, 157. doi:10.1186/1471-2164-9-157
- Kozomara, A., and Griffiths-Jones, S. (2013). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids Res.* 42 (D1), D68–D73. doi:10.1093/nar/gkt1181
- Kumar, P., Anaya, J., Mudunuri, S. B., and Dutta, A. (2014). Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *BMC Biol.* 12, 78. doi:10.1186/s12915-014-0078-0
- Kumar, P., Mudunuri, S. B., Anaya, J., and Dutta, A. (2015). tRFdb: a database for transfer RNA fragments. *Nucleic Acids Res.* 43 (Database issue), D141–D145. doi:10.1093/nar/gku1138
- Kuscu, C., Kumar, P., Kiran, M., Su, Z., Malik, A., and Dutta, A. (2018). tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. *RNA* 24 (8), 1093–1105. doi:10.1261/rna.066126.118
- Langenberger, D., Bermudez-Santana, C., Hertel, J., Hoffmann, S., Khaitovich, P., and Stadler, P. F. (2009). Evidence for human microRNA-offset RNAs in small RNA sequencing data. *Bioinformatics* 25 (18), 2298–2301. doi:10.1093/bioinformatics/btp419
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. methods* 9 (4), 357–359. doi:10.1038/nmeth.1923

- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10 (3), R25. doi:10.1186/gb-2009-10-3-r25
- Lee, Y. S., Shibata, Y., Malhotra, A., and Dutta, A. (2009). A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev.* 23 (22), 2639–2649. doi:10.1101/gad.1837609
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi:10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp324
- Lin, H., Huang, Z. P., Liu, J., Qiu, Y., Tao, Y. P., Wang, M. C., et al. (2018). MiR-494-3p promotes PI3K/AKT pathway hyperactivation and human hepatocellular carcinoma progression by targeting PTEN. *Sci. Rep.* 8 (1), 10461. doi:10.1038/s41598-018-28519-2
- Lopes, M., Louzada, S., Ferreira, D., ssimo, G., rio, D., Gama-Carvalho, M., et al. (2023). Human Satellite 1A analysis provides evidence of pericentromeric transcription. *BMC Biol.* 21 (1), 28. doi:10.1186/s12915-023-01521-5
- Louzada, S., Lopes, M., Ferreira, D., Atega, F., Escudeiro, A., Gama-Carvalho, M., et al. (2020). Decoding the role of satellite DNA in genome architecture and plasticity—an evolutionary and clinical affair. *Genes* 11 (1), 72. doi:10.3390/genes11010072
- Lyons, S. M., Fay, M. M., and Ivanov, P. (2018). The role of RNA modifications in the regulation of tRNA cleavage. *FEBS Lett.* 592 (17), 2828–2844. doi:10.1002/1873-3468.13205
- Miesen, P., Ivens, A., Buck, A. H., and van Rij, R. P. (2016). Small RNA profiling in dengue virus 2-infected Aedes Mosquito cells reveals viral piRNAs and novel host miRNAs. *PLoS Negl. Trop. Dis.* 10 (2), e0004452. doi:10.1371/journal.pntd.0004452
- Moncini, S., Salvi, A., Zuccotti, P., Viero, G., Quattrone, A., Barlati, S., et al. (2011). The role of miR-103 and miR-107 in regulation of CDK5R1 expression and in cellular migration. *PLoS ONE* 6 (5), e20038. doi:10.1371/journal.pone.00220038
- Nunes, A., Ribeiro, D. R., Marques, M., Santos, M. A. S., Ribeiro, D., and Soares, A. R. (2020). Emerging roles of tRNAs in RNA virus infections. *Trends Biochem. Sci.* 45 (9), 794–805. doi:10.1016/j.tibs.2020.05.007
- Persson, H., Kvist, A., Vallon-Christersson, J., Medstrand, P., Borg, A., and Rovira, C. (2009). The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. *Nat. Cell Biol.* 11, 1268–1271. doi:10.1038/ncb1972
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26 (6), 841–842. doi:10.1093/bioinformatics/btq033
- Rashad, S., Niizuma, K., and Tominaga, T. (2020). tRNA cleavage: a new insight. *Neural Regen. Res.* 15 (1), 47–52. doi:10.4103/1673-5374.264447
- Roa-Linares, J. C. G. G. (2019). The loss of function of Cyclin-Dependent Kinase 5 (CDK5) alters the Cytoskeleton and decrease the *in vitro* Dengue Virus-2 infection. *Acta Biol. Colomb.* 474. doi:10.15446/abc.v24n3.79347
- Saraiya, A. A., and Wang, C. C. (2008). snoRNA, a novel precursor of microRNA in *Giardia lamblia*. *PLoS Pathog.* 4, e1000224. doi:10.1371/journal.ppat.1000224
- Schopman, N. C., Heynen, S., Haasnoot, J., and Berkhout, B. (2010). A miRNA-tRNA mix-up: tRNA origin of proposed miRNA. *RNA Biol.* 7 (5), 573–576. doi:10.4161/rna.7.5.13141
- Selitsky, S. R., Baran-Gale, J., Honda, M., Yamane, D., Masaki, T., Fannin, E. E., et al. (2015). Small tRNA-derived RNAs are increased and more abundant than microRNAs in chronic hepatitis B and C. *Sci. Rep.* 5 (1), 7675. doi:10.1038/srep07675
- Shi, W., Hendrix, D., Levine, M., and Haley, B. (2009). A distinct class of small RNAs arises from pre-miRNA-proximal regions in a simple chordate. *Nat. Struct. Mol. Biol.* 16, 183–189. doi:10.1038/nsmb.1536
- Stadler, P. F., Chen, J. J., Hackermuller, J., Hoffmann, S., Horn, F., Khaitovich, P., et al. (2009). Evolution of vault RNAs. *Mol. Biol. Evol.* 26, 1975–1991. doi:10.1093/molbev/msp112
- Taft, R. J., Glazov, E. A., Lassmann, T., Hayashizaki, Y., Carninci, P., and Mattick, J. S. (2009). Small RNAs derived from snoRNAs. *RNA* 15, 1233–1240. doi:10.1261/rna.1528909
- Tam, S., Tsao, M. S., and McPherson, J. D. (2015). Optimization of miRNA-seq data preprocessing. *Briefings Bioinforma.* 16 (6), 950–963. doi:10.1093/bib/bbv019
- Taxis, T. M., Kehrl, M. E., D'Orey-Branco, R., and Casas, E. (2018). Association of transfer RNA fragments in white blood cells with antibody response to bovine leukemia virus in Holstein cattle. *Front. Genet.* 9, 236. doi:10.3389/fgene.2018.00236
- Telonis, A. G., Kirino, Y., and Rigoutsos, I. (2015). Mitochondrial tRNA-lookalikes in nuclear chromosomes: could they be functional? *RNA Biol.* 12 (4), 375–380. doi:10.1080/15476286.2015.1017239
- tRFdb (2020). A relational database of Transfer RNA related Fragments. Available from <http://genome.bioch.virginia.edu/trfdb/index.ph>.
- Tycowski, K. T., Guo, Y. E., Lee, N., Moss, W. N., Vallery, T. K., Xie, M., et al. (2015). Viral noncoding RNAs: more surprises. *Genes & Dev.* 29 (6), 567–584. doi:10.1101/gad.259077.115
- Velandia-Huerto, C. A., Berkemer, S. J., Hoffmann, A., Retzlaff, N., Romero Marroquín, L. C., Hernandez-Rosales, M., et al. (2016). Orthologs, turn-over, and remodeling of tRNAs in primates and fruit flies. *BMC Genomics* 17 (1), 617. doi:10.1186/s12864-016-2927-4
- Wang, Q., Lee, I., Ren, J., Ajay, S. S., Lee, Y. S., and Bao, X. (2013). Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection. *Mol. Ther.* 21 (2), 368–379. doi:10.1038/mt.2012.237
- Zhou, J., Liu, S., Chen, Y., Fu, Y., Silver, A. J., Hill, M. S., et al. (2017). Identification of two novel functional tRNA-derived fragments induced in response to respiratory syncytial virus infection. *J. Gen. Virol.* 98 (7), 1600–1610. doi:10.1099/jgv.0.000852
- Zhu, P., Yu, J., and Zhou, P. (2020). Role of tRNA-derived fragments in cancer: novel diagnostic and therapeutic targets tRFs in cancer. *Am. J. Cancer Res.* 10 (2), 393–402.